Metrics

Introduction

Metrics are used in most areas of business and government, from measuring hospital mortality to call centre response times. They have been common within pharmaceutical industry clinical trials for some years and have been used for internal company performance and as part of agreements with external suppliers, such as Contract Research Organisations (CROs). Many of the metrics relate to site performance, such as recruitment rate, but they are also relevant for the other processes, including statistics and programming. However there are no readily available published metrics that relate specifically to the statistical and programming activities in clinical trials (for internal company performance or as part of agreements with external suppliers).

The Pharmaceutical Contract Management Group (PCMG) and European CRO Federation (EUCROF) set up a Metrics Working Group. This group has initially looked at clinical CRO operational metrics but wanted to also consider biometrics metrics, in the broader context of assessing performance of both sponsor and CRO. Therefore PCMG/EUCROF and PSI set up a small working group to consider biometrics metrics and the results are summarised below.

PCMG

Founded in 1994 PCMG has evolved into an international organisation for pharmaceutical company professionals who manage the outsourcing of clinical development activities. The mission of this independent, not-for-profit, association is to establish, continuously improve and openly share best practices in clinical outsourcing. PCMG together with EUCROF established a CRO metrics working group in 2014 and presented the outputs at the EUCROF conference in February 2015 and the PCMG annual conference in June 2015. Together with PSI and others, PCMG are developing metrics in a variety of functional areas for the general benefit of our industry.

Process

The project was instigated by PCMG and EUCROF, and a PCMG representative contacted the PSI Chair for PSI's support, which was given. The joint chairs of the group were identified as Lan Bandara for PCMG and Ray Harris for PSI. A request for volunteers to join the group was publicised by PCMG, EUCROF and put in the PSI eNews, with an explanation of the purpose and expected activities. The contributing volunteers from these requests, listed below, joined the group. The volunteers were from different parts of the pharmaceutical industry and CROs and different locations and they contributed their knowledge and experience to the development of the metrics. The goal of the group was to develop metrics that relate specifically to the statistical and programming activities of both sponsors and CROs in clinical trials.

The group met by teleconference a number of times, with additional email communication, to discuss and agree suitable metrics, key considerations, and, finally, this publication.

The Metrics

The metrics presented in this publication (Table 1) are not intended to be exhaustive but do include many key areas. There are, of course, many others that could be measured and may have specific relevance. There is one metric identified that relates to cost, all the others are either time or quality based, though these are not exclusive categories and there are strong associations between quality, timelines and cost (direct, or indirect through resource allocation).

It should be noted that all but one of the metrics could be used for biometrics performance regardless of who is providing the support, the sponsor or a CRO.

Whilst Table 1 gives all the metrics that were agreed upon as relevant; there were other metrics discussed by the working group, drawing on the wide range of experience in the group.

The quality metrics reference errors or changes. For a metric to accurately represent its measure it is important to understand what affects the measure. For instance, it is a very different issue if an error in an output after database lock is a misspelling in a title compared to an adverse event miscoded in the database. Where quality metrics are used when outsourcing biometrics it is important that the metrics are understood and agreed by all the parties, particularly what they measure, and when calculating them what is the potential root cause of any out-of-specification results.

The metrics are discussed in more detail below.

Statistical Analysis Plan (SAP)

The first three metrics relate to the SAP (the TLG templates may be part of the SAP or a separate deliverable but the metrics are the same). The final protocol is a clear milestone and therefore the time from this to the SAP approval is an obvious metric. The expectation is that the SAP will be ready as soon as possible, as it is used in discussions with the regulators, and enables the preparation of programming for the reporting of the study. Also, it is convenient and efficient for the study team to move on from developing the protocol to providing input into, and review of, the SAP. The first SAP after the protocol finalisation is version 1.0 but often it is felt useful to call it the final version, as is the protocol, while being aware that there may need to be amendments later, due to, for instance, protocol amendments.

Outputs

Most of the relevant metrics relate to the statistical analysis outputs (metrics 4, 6, 7, 8, 9 and 10).

For most phase II and III studies a dry run of the outputs will be planned. This is to ensure that the programming is in place and the outputs that are produced match the agreed templates. Frequently more than one dry run will be produced. The error rate on the last is expected to be zero. Outputs are generally validated and with quality checks passed. Errors may be defined as critical or non-critical with different thresholds (e.g. 0% and 10% respectively).

The metrics use database lock as a start time, however there will be a number of steps, such as unblinding and creation of ADaM datasets, before programming and analysis can start.

This needs to be considered when applying these metrics since the activities may not be within the full remit of Biostatistics and Programming.

It was agreed that once the database has been locked, there should be no errors in any of the outputs produced or in the ADaM datasets (metric 5).

The time, in working days, from database lock to top line outputs was much discussed. Though everyone agreed that top line outputs were critical there was a range of experience of how many outputs were expected. Most felt that more than 20 outputs would be excessive, though more than 50 have been seen. Less than about 10 outputs, provided in about a working week, was not unrealistic.

It is common for the full study outputs to be delivered in batches, with the most important prioritised. The reporting metric will usually relate to the completion of the statistical reporting with the final batch, and the time to that will depend on the study complexity and the agreed delivery and review process.

Other metrics

There was discussion about the requirement to have a meeting before database lock to agree populations, patient by patient, resolving any outstanding issues. In the end it was felt that this was not a true metric, though it was agreed that it was a critical step.

Only metric 11, time to program outputs or datasets, explicitly measures productivity. No consensus was reached on targets for this. However, this is, perhaps, not surprising given the different approaches to programming efficiency including wide use of macros and reporting systems. The quality control may be different depending on the output and process used. So what seems like an easy question does not have an easy answer.

Metric 12 is the only one that directly relates to cost. Cost is linked to analysis specifications, and agreed between the sponsor and the CRO.

Discussion

This process has provided a useful set of metrics that statistical and programming representatives from a range of organisations in the pharmaceutical industry and CROs have discussed and agreed upon. The targets for many of the metrics were non-controversial and easy to agree. For others agreement was more difficult. This is not surprising as different organisations will set different priorities on the stages of the clinical trials and also have different approaches to ensuring high quality timely completion. This should be born in mind when comparing these targets with your own organisation's targets. However, the biometrics metrics proposed here can serve as guidance to healthcare professionals who are involved with resourcing and managing of biostatistics and programming services and to improve internal processes.

Finally, the old adage, the balance of cost, quality and timeliness applies within biometrics as much as to other functions. For example, reduced timelines can be achieved by taking activities off the critical path and implementing increased parallel reviews but this in turn increases costs.

Working Group Membership

Chairs:
Lan Bandara, Eisai, UK for PCMG
Ray Harris, Eisai, UK for PSI

Members:
Federica Alessi, BMS, Belgium
Mikaël Le Bouter, Arlenda, Belgium
Antonio Cagnazzo, CROSNT, UK
David Inman, Daiichi Sankyo, UK
Milena Kurtinecz, GSK, USA
Caroline Morgan, Cytel, UK
Alan Phillips, ICON, UK
Yannis Planoudis, Nestec, Switzerland

No	Metric Type	Description	Calculation	Target	Comments
1	Timeline	Time from final protocol to statistical analysis plan / TLG templates version 1.0 (final).	Number of working days.	Time is agreed on a study basis.	Version 1.0 is the version that enables programming to start and is provided to regulators if required.
2	Quality	Number of review iterations of analysis plan, TLG templates and ADaM specifications.		Two review cycles	This may reflect errors or changes by Biostats, Clinical and others.
3	Quality	Changes in analysis plan and TLG templates following 'final' version, other than due to protocol amendments or regulatory feed-back.		No changes	Differentiate between errors and changes.
4	Quality	Error rate in last dry run output (unless only one dry run planned)	Numbered outputs with any error as a per cent of total numbered outputs.	0%	Errors are any mismatches between the dry run and the documented specifications.
5	Quality	Per cent of ADaM dataset variables with errors after database lock.		0%	
6	Timeline	Time from database lock to top-line results as per the SAP.		Number of outputs and working days to deliver: 1-5: 3-5 Days 6-10: 5-7 Days 11-20: 5-10 Days >20: >10 Days	Responsibilities for steps from database lock to availability of datasets for reporting may differ in different companies. It is very rare for a top-line to have >20 outputs.
7	Quality	Error rate in output following database lock: top-line	Numbered outputs with any error as a per cent of total numbered outputs	0%	Errors include typos, programming errors and data errors, the later not necessarily attributable to Biostatistics and Programming.
8	Timeline	Time taken from database lock to draft outputs.	Working days	12-30 days for final batch.	Time will be on a study by study basis and be related to complexity. Metric may be to first or last batch.

9	Quality	Error rate in output following database lock: draft outputs	Numbered outputs with any error as a per cent of total numbered outputs.	0%	Errors include typos, programming errors and data errors, the later not necessarily attributable to Biostatistics and Programming.
10	Quality	Error rate in output following database lock: final outputs	Numbered outputs with any error as a per cent of total numbered outputs	0%	Errors include typos, programming errors and data errors.
11	Timeline	Average time taken to program unique table, listing and graph, and the same for each repeat: time for standard and complex ADaM datasets.		None agreed	Includes programing and QC.
12	Costs	Actual cost vs. originally agreed budget.		<20% excess	Limit for additional costs depends on study complexity.